

Permanent City Research Online URL: http://openaccess.city.ac.uk/1082/

Copyright & reuse
City University London has developed City Research Online so that its users may access the research outputs of City University London's staff. Copyright © and Moral Rights for this paper are retained by the individual author(s) and/ or other copyright holders. Users may download and/ or print one copy of any article(s) in City Research Online to facilitate their private study or for non-commercial research. Users may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. All material in City Research Online is checked for eligibility for copyright before being made available in the live archive. URLs from City Research Online may be freely distributed and linked to from other web pages.

Versions of research
The version in City Research Online may differ from the final published version. Users are advised to check the Permanent City Research Online URL above for the status of the paper.

Enquiries
If you have any enquiries about any aspect of City Research Online, or if you wish to make contact with the author(s) of this paper, please email the team at publications@city.ac.uk.
• Who?
• Where?
• What?
Why did we undertake this research?

- Social media are an important source of information during a crisis.
- Humanitarian organisations are reluctant to use this information in the response effort because untrustworthy and inaccurate information can cost lives.
- Organisations such as Ushahidi use crowdsourcing to address these concerns, but crowdsourcing introduces further uncertainty.
- We’re interested in evaluating the uncertainty, and the potential bias, in crowdsourced crisis information.
- Crowdsourced crisis information is highly geographic; there are several characteristics of uncertainty relevant to geographic information (including trust and accuracy).
- We started with accuracy.
1. What types of locality descriptions are present in crowdsourced crisis information? (i.e. classification)

2. Are the proportions of these types different to those present in related datasets? (i.e. comparison)
• We classified locality descriptions in about 3,600 incident reports related to the 2010 earthquake in Haiti.

• We used a classification that was developed alongside a georeferencing method (the point-radius method). Why?
  – The method was developed to georeference locality descriptions.
  – It provides an estimate of the uncertainty associated with the georeferencing process.
  – It has been applied to a variety of related datasets:
    ▶ Records of artefacts in natural history collections; these records are stored in MaNIS (the Mammal Networked Information System).
    ▶ Historical records of search and rescue incidents.

• We compared the Haiti and MaNIS datasets.
How did we address our research questions?

- Two papers discuss the categories of locality descriptions in the MaNIS dataset. The categories they identify are slightly different. We combined them. Our combined classification has 12 categories.

- Three participants independently classified the locality descriptions in the Haiti dataset. (The classification process took about 4 hours for each participant to complete.)

- The order of the locality descriptions was randomised for each participant.

- Each participant was guided by the definitions in the table and examples from the MaNIS dataset.

<table>
<thead>
<tr>
<th>Code</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>Unsure</td>
</tr>
<tr>
<td>C</td>
<td>Coordinates</td>
</tr>
<tr>
<td>F</td>
<td>Feature</td>
</tr>
<tr>
<td>P</td>
<td>Path</td>
</tr>
<tr>
<td>J</td>
<td>Junction</td>
</tr>
<tr>
<td>FOH</td>
<td>Offset from a feature or path at a heading</td>
</tr>
<tr>
<td>NF</td>
<td>Near a feature or path</td>
</tr>
<tr>
<td>FS</td>
<td>Subdivision of a feature or path</td>
</tr>
<tr>
<td>FDO</td>
<td>Orthogonal offsets from a feature</td>
</tr>
<tr>
<td>FH</td>
<td>Heading from a feature, no offset</td>
</tr>
<tr>
<td>FO</td>
<td>Offset from a feature or path, no heading</td>
</tr>
<tr>
<td>BF</td>
<td>Between features or paths</td>
</tr>
</tbody>
</table>

Table: Combined classification of locality descriptions
Here we see the frequency of locality descriptions in each category, by each participant, for the Haiti dataset.

- ‘Feature’ is 1st for all participants.
- ‘Path’ is 2nd for P1 and P2, 3rd for P3.
- ‘Unsure’ is 2nd for P3, 3rd for P1 and 5th for P2.

Using Fleiss’ kappa, we tested the degree to which the observed amount of agreement between the participants exceeded what would be expected if each participant were to categorise each locality description at random.

- $\kappa = 0.42$ (‘moderate agreement’)
What did we find?

Method:

- We classified partial agreement cases by simple majority vote.
- We classified disagreement cases as ‘Uncertain’.

Here we see the frequency of locality descriptions in each category, for the Haiti dataset.

- Most cases (about 71%) are ‘Feature’.
- 419 cases (about 12%) are ‘Uncertain’.
What did we find?

Method:

- We removed all ‘Uncertain’ and ‘Coordinates’ cases (about 13% of the Haiti dataset) to allow a like-for-like comparison with the MaNIS dataset.

Here we see the proportion of locality descriptions in each category, for the MaNIS and Haiti datasets.

- Like the MaNIS dataset, most locality descriptions in the Haiti dataset describe features. However, the proportions are very different (MaNIS about 51%; Haiti about 82%).

- There are similar proportions of paths (about 9%) and junctions (about 1%).
What did we find?

Here we see the rank and proportion of locality descriptions in each category, for the MaNIS and Haiti datasets, ordered by the Haiti dataset.

- 0.5% of locality descriptions in the Haiti dataset contain offsets and headings (FH, FOH, FO, FOO—the lower four rows). The figure is 27.0% for the MaNIS dataset.

- 91.0% of locality descriptions in the Haiti dataset are either features or paths (F, P—the upper two rows). The figure is about 60% for the MaNIS dataset.

<table>
<thead>
<tr>
<th>Code</th>
<th>MaNIS (%)</th>
<th>Haiti (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>NF</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>FS</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>J</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>BF</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>FH</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>FOH</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>FO</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>FOO</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Table: Category rank and proportion, MaNIS and Haiti datasets
How did we interpret what we found?

- Location is seldom described using offsets and headings in the Haiti dataset. (FH, FOH, FO, FOO)
- Location is often described in terms of features (such as named places) and paths (such as roads). (F, P)
This finding appears to conform previous research: ‘nature’ and ‘phase’ affect the information about an emergency event. For example, during the impact and recovery phases, information from social media contains a high proportion of GI, and this GI relates to well-defined geographic objects (Vieweg et al., 2010). This finding might generalise!

The uncertainty associated with georeferencing features and paths should be less than that associated with offsets and headings, as they have to be offsets and headings from somewhere.

If we were to georeference the locality descriptions in the Haiti dataset, we could expect reasonably accurate results. We could compare these results to the locations produced by crowdsourcing to evaluate one aspect of the uncertainty in crowdsourced crisis information.
How did we interpret what we found?

- Remember *Fleiss' kappa*? There was ‘moderate agreement’ between the participants. (For example, about 12% of cases in the Haiti dataset are ‘Uncertain’.) In other words, *ambiguity* presents a significant challenge.

- About 5% of the Haiti dataset are ‘Near a feature or path’. So, there is also *vagueness*. Anecdotally, there are many *vague places* such as references to IDP (Internally Displaced Person) camps (*vernacular geography* and *naive geography*).

- We didn’t investigate the *precision* (resolution, scale) of locality descriptions.
- Locality descriptions in the Haiti dataset favour more certain locations (e.g. features, paths) rather than less certain locations (e.g. that contain offsets and headings).

- Georeferencing these locality descriptions could provide a basis for comparison with the locations produced by crowdsourcing.

- But ambiguity and vagueness present significant challenges: It’s complex!
Future work

We plan to:

• use alternative sources of information (e.g. OpenStreetMap) to reduce ambiguity and vagueness, and explore precision;

• explore related datasets (e.g. Libya);

• develop a geovisualization tool for exploration and analysis.
We would like to thank:

- Roger Beecham and Sarah Goodwin;
- those behind the Haiti Crisis Map;
- the anonymous reviewers.